2015年6月19日　統計数理研究所　オープンハウス

外れ値の割合をも推定するロバスト推定

藤澤　洋徳　数理・推論研究系　教授

はじめに

Fujisawa and Eguchi (2008) では，外れ値の割合が大きくても，バイアスを十分に小さくできるロバスト推定を考えた．そこでは，γ-divergence が，主要な枠割を果たす．ある種のピタゴリアン関係も成り立ち，パラメータ推定の自然さも見てとれる．代表的な分布の幾つかでは，簡単なパラメータ推定アルゴリズムを，きれいな形で提案できる．そして，外れ値の割合が大きくてもバイアスを十分に小さくできるロバスト推定を可能にするダイバージェンスは，ある意味では γ-divergence に限ることも示した．

その結果においては，外れ値の分布を自動的に無視する機構と，目的分布の割合を自動的に調整する機構が，本質であった．本研究では，目的分布の割合（言い換えれば外れ値の割合）をも推定できるように，拡張されたモデルを用意することで，γ-divergence に限らずに，上記の目的を可能にすることを考えた．なお，その際には，やはり，γ-divergence が重要な枠割を担っていることも，見てとれる．

相互エントロピーとダイバージェンス

p, q: $p(x) \geq 0$. $q(x) \geq 0$.　（普通は密度関数だけに制限される.）

相互エントロピー $d(p,q)$:

$$d(p,q) \geq d(p,p).$$

等号は$p=q$のときのみに成り立つ．

ダイバージェンス $D(p,q)$:

$$D(p,q) = d(p,q) - d(p,p) \geq 0$$

パラメータ推定:

$$\hat{\theta} = \arg\min_{\theta} D(\bar{p}, q_\theta) = \arg\min_{\theta} d(\bar{p}, q_\theta)$$

$\bar{p}$: 経験密度関数など

注：最尤推定はKLダイバージェンスと対応する．

拡張: 等号が成り立つのは，pとqが密度関数に制限された場合においては，$p=q$のときに限る．

注：γ-divergence は，通常のダイバージェンスではないが，この意味での拡張されたダイバージェンスではある．この拡張は，外れ値の割合が大きい場合にもバイアスのないパラメータ推定を行うときに，有効となる (Fujisawa and Eguchi, 2008).

Hölderダイバージェンス

$$d_H(p,q) = \phi\left(\int p(x)q(x)^\gamma dx \Big/ \int q(x)^{1+\gamma}\right) \int q(x)^{1+\gamma} dx$$

$$\gamma > 0 \qquad \phi(1) = -1 \qquad \phi(z) \geq -z^{1+\gamma} \quad (z \geq 0)$$

注：経験推定可能である相互エントロピーのあるクラスにおいて，ある種のアフィン不変性をもつものは，適当な仮定の下では，上記の相互エントロピーに限る (Kanamori and Fujisawa, 2014).

例：

γ-相互エントロピー: $\phi(z) = -z^{1+\gamma}$　（これは下限）

β-相互エントロピー: $\phi(z) = \beta - (1+\beta)z$

良いロバスト推定とは

データ発生分布：

$$g(x) = (1-\varepsilon)f_*(x) + \varepsilon\delta(x).$$

$f_*(x)$: 目的分布
$\delta(x)$: 外れ値に対応する分布
ε: 外れ値の割合

普通の推定の問題点：普通は，パラメータを推定するときに，データ発生分布gとモデルf_θとのダイバージェンスを最小化しようとする．外れ値の影響がないということは，その項が推定において自動的に（ほぼ）無視できるのであると考えよう．この考えは次の数式で表現できる：

$$\begin{aligned}\hat{\theta} &= \arg\min D(g, f_\theta) \\ &= \arg\min D\big((1-\varepsilon)f_*(x) + \varepsilon\delta(x), f_\theta\big) \\ &\approx \arg\min D\big((1-\varepsilon)f_*(x), f_\theta\big)\end{aligned}$$

通常のダイバージェンスでは，大きさが違う二つの関数$(1-\varepsilon)f_*$とf_θを十分に近似できない．ここに問題が生じる．

γ-divergence　の効用：拡張されたダイバージェンスである　γ-divergence を使うと，この定数$(1-\varepsilon)$を自動的に（ほぼ）無視できる (Fujisawa and Eguhi, 2008).

拡張モデル

本研究では，定数$(1-\varepsilon)$を自動的に無視するのではなく，拡張されたモデルを考えることで，その定数をも同定することを考える．

$$m_\eta(x) = \xi f_\theta(x), \qquad \eta = (\xi, \theta).$$

定理　密度関数$g(x)$と拡張モデル$m_\eta(x)$との Hölder divergence は

$$\xi = \int g(x)f_\theta(x)^\gamma dx \Big/ \int f_\theta(x)^{1+\gamma} dx$$

のときに最小化される．結果として次も成り立つ：

$$\min_{\xi} d_H(g, m_\eta) = -\exp\{-\gamma(1+\gamma)d_\gamma(g, f_\theta)\}.$$

注：どのような Hölder divergence であったとしても，拡張モデル$m_\eta(x) = \xi f_\theta(x)$を考えることで，モデルパラメータ$\theta$の推定においては γ-divergence の最小化と同等のことさえ考えればよい．

補助定理　汚染分布と目的分布の間に次の仮定が成り立つとする：$\int \delta(x) f_*^\gamma(x)dx \approx 0$. 適当な仮定の下で以下が成り立つ：

$$\hat{\xi} = \arg\min_{\xi} d_H(g, m_\eta) \approx 1 - \varepsilon.$$

参考文献

Fujisawa, H. and Eguchi, S. (2008). Robust parameter estimation with a small bias against heavy contamination. Journal of Multivariate Analysis, Vol.99, 2053-2081.

Kanamori, T. and Fujisawa, H. (2014). Affine invariant divergences associated with composite scoring rules and their applications. Bernoulli, Vol.29, 2278-2304.

Kanamori, T. and Fujisawa, H. (2015). Robust estimation under heavy contamination using unnormalized models. Biometrika (in press).